Potential Energy Surfaces and Conformational Transitions in Biomolecules: A Successive Confinement Approach Applied to a Solvated Tetrapeptide

Sergei V. Krivov,[1,2] Sergei F. Chekmarev,[1,3] and Martin Karplus[2,4]

[1]*Department of Physics, Novosibirsk State University, 630090 Novosibirsk, Russia*
[2]*Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France*
[3]*Institute of Thermophysics, 630090 Novosibirsk, Russia*
[4]*Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138*
(Received 15 February 2001; published 2 January 2002)

A simple approach for the efficient exploration of the potential energy surface of a many-body system is presented. The method uses Langevin dynamics trajectories that are successively confined in the various basins of the potential energy surface. The approach is illustrated by determining the potential energy surface, and the thermodynamic and kinetic properties of a solvated model for the alanine tetrapeptide, the shortest peptide that can form an α-helical turn. All possible *cis* isomers are sampled, even though the barriers separating them are as high as 25 kcal/mole. Comparisons with conventional Langevin dynamics confirm the greater efficacy of the approach.

DOI: 10.1103/PhysRevLett.88.038101 PACS numbers: 87.15.–v

The potential energy surfaces of mesoscopic systems (systems with many degrees of freedom that are in the 10 to 1000 Å range) are fundamental to an understanding of their structural, thermodynamic, and dynamic properties. Of particular interest are mesoscopic systems of importance in biology, such as proteins, nucleic acids, and lipid membranes. The complexity of the behavior of such systems has been demonstrated for proteins in the native state, as exemplified by myoglobin [1,2] and bacteriorhodopsin [3], and for folding to the native state [4].

An approach for describing such complex potential energy surfaces (PES) based on topological mapping via disconnectivity graphs has been introduced recently [5]. The disconnectivity graph shows which minima are connected by pathways lying below a certain energy threshold. They have been applied to classify "archetypical energy landscapes" for Lennard-Jones clusters, buckministerfullarene, and a model water cluster [6]. The essential element in the construction of the disconnectivity graph is a knowledge of the local minima and the saddles connecting them. Since determination of a meaningful portion of the PES becomes more difficult as the size of the system increases, it is essential to develop sampling methods that go beyond the widely used Monte Carlo (MC) [7] or molecular dynamics (MD) approaches [8].

Most of the complex systems of interest have a PES that consists of multiple low energy regions (basins) separated by barriers that are large with respect to kT; a schematic PES of this type in one dimension is shown in Fig. 1. The behavior of a system with such a PES is often nonergodic at normal temperatures on the time scales accessible to standard MD or MC methods on current computers; i.e., the sampled distribution depends on the starting structure. Thus, for effective sampling, it can be useful to bias the search algorithm so that it violates the equilibrium Boltzmann distribution corresponding to the PES. A number of methods have been proposed for this purpose [9]. Most of the methods in [9] are useful for obtaining equilibrium properties. They are of limited value for a detailed survey of the PES because they miss important low enthalpy, low entropy minima. In this Letter, we present an alternative approach and illustrate it by a full exploration of the PES of a solvated tetrapeptide (Fig. 2), the shortest peptide that can form an α-helical turn. It has a nontrivial PES, but still is in the reach of ordinary MD, so that we can determine the speedup of the new method. The peptide is represented by an empirical force field [10] with an implicit solvent model that has been shown to be accurate for peptides [11]. Since a very similar tetrapeptide has been analyzed *in vacuo* [5,12], the comparison is of considerable interest. The power of the method is demonstrated by the fact that not only the manifold of states associated with the lowest energy all-*trans* peptide bonded isomers are well sampled, but that all the *cis* peptide isomers (mono *cis* through tetra *cis*), which are separated by high barriers from the all-*trans* form, are found without explicitly searching them. We demonstrate that the equilibrium thermodynamics of the basins and the kinetics of the transitions between them are determined efficiently, as well.

The present approach makes use of Langevin dynamics (LD) to generate the trajectory of the system (though Newtonian MD [13] or MC could be used instead) and is based on a very simple concept that gives it a wide range of

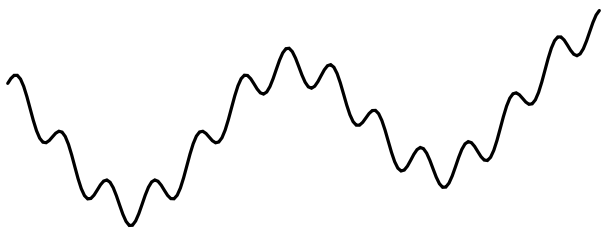

FIG. 1. A model one-dimensional potential surface (see text).

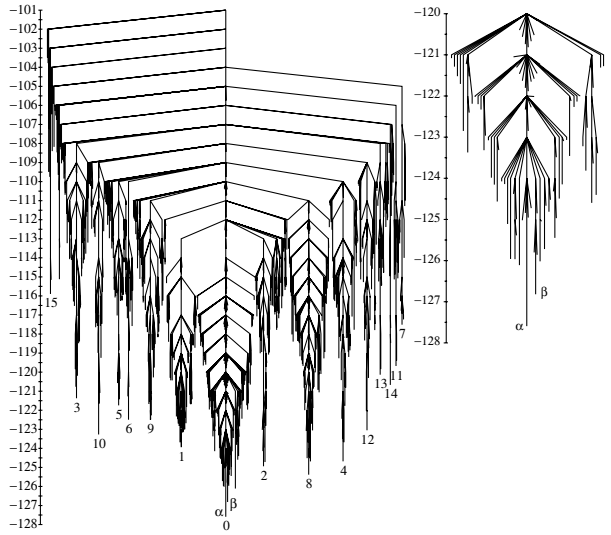

FIG. 2. Coarse graining disconnectivity graph of the entire PES of the tetrapeptide; energy spacing is in kcal/mol. The different *cis* isomers are numbered according to a reverse binary rule (i.e., the count is made from left to right) with 0 and 1 standing for the *trans* and *cis* isomer of the peptide group, respectively. For example, 7 = (1110) is the *cis-cis-cis-trans* conformation. α and β label the α-helix and β-strand (extended) conformers, respectively. The right panel shows the magnified view of the low energy all-*trans* part of the PES.

applicability. The essential idea is that the system is confined successively to different basins on the PES, with the choice of basins for detailed examination, after they have been visited at least once, and the length of time spent in the basins being determined *by the problem under investigation* (in other approaches [9] the system still follows inherent dynamics on a deformed PES). It offers essentially unlimited flexibility in the distribution of the residence times of the system in different regions of the PES and provides a general framework for constructing optimal simulation schemes.

To indicate why exploration of the PES by successive confinement is likely to be significantly more efficient than conventional MD or MC, we consider the one-dimensional model surface depicted in Fig. 1. It consists of two clusters of basins ("superbasins"), and it is assumed for simplicity that all the basins are identical except for their minimum energies. In particular, they have the same mean lifetime τ, and the same probability, p, for the system to go to the neighboring higher energy basin in a given time (the probability to go in the reverse direction is $1 - p$) with $p \ll 1$; it is for this case that the confinement method is expected to be most useful. The system is placed at the bottom of one of the superbasins. If conventional MD at a given T is used, the system freely goes up and down among the basins and the mean time for the system to reach the nth basin from the bottom of the first superbasin is $t_{\text{conv}} = \tau/p^n$. For the Fig. 1 surface, if the superbasin barrier is much greater than kT, the system is likely to remain in the original superbasin with conventional dynamics. In the successive confinement method, the system is not allowed to return to a lower energy basin because it has been sampled already. The mean time for the system to go to the next higher energy basin is τ/p [i.e., it is the same as for t_{conv}; the mean time to go to the neighboring lower energy basin is $\tau/(1 - p) \sim \tau$] and the time to reach the nth basin is equal to $t_{\text{conf}} = n\tau/p$. The gain in efficiency, if the same path is followed in the conventional and confinement simulations, is equal to the ratio $t_{\text{conv}}/t_{\text{conf}} = p^{1-n}/n$. For $n = 4$ as in Fig. 1, and $p = 0.1$, for example, we find $t_{\text{conv}}/t_{\text{conf}} = 250$. In the tetrapeptide simulation, a comparison of the two approaches can be made for the transition from the extended (β) strand to the α-helix (see below). In the conventional simulation, the fastest pathway involves three transitions, while it requires five in the confinement simulation; in both cases p is about 0.07. Using the expressions for t_{conv} and t_{conf}, we obtain $t_{\text{conv}}/t_{\text{conf}} = p^{-2}/5 = 41$.

With Langevin dynamics, the procedure is as follows. The system is placed in a certain (ith) basin and a MD run is begun at a temperature T; both the physical temperature or a higher temperature can be used. At regular intervals the system is quenched (to 0 K) to check if the trajectory is still in the given basin or has left it; for this, the convergence of the quenched structures to or divergence from that corresponding to the minimum of the basin is examined [2,14]. At the current quench, atomic coordinates at the phase point ($\mathbf{x}_\alpha$), where α indicates the basin, are stored; if the system is in the original (ith) basin, $\alpha = i$, the MD run is continued. If the system has left the ith basin, the system is placed back into the ith basin (at the point $\mathbf{x}_i$ associated with the previous quenching), and a new trajectory is initiated with velocities chosen at random from a Maxwellian distribution at the temperature of the simulation. In this way, the system can be kept in the given basin for any desired time. Since all visited basins are recorded, we can calculate not only the equilibrium properties corresponding to the given (ith) basin, but also determine the probabilities to pass into the connected basins. Once the properties of interest associated with basin i have converged, the system is allowed to pass into neighboring basins of interest and the described procedure is repeated.

The choice of the next basins to be visited depends on the goal of the study. For systems that are small enough so that a sampling of all (or a very large part) of the basins on the PES is possible, one can allow the system to pass into each new basin that has not been previously sampled and study it in turn. However, if the system is large, the goal may be to survey just a part of the PES (e.g., the low energy region that is important at ordinary temperatures), or to obtain a "coarse" survey of the entire PES. In the first case, preference is given to the basins which are in the portion of the PES of interest; in the second case, the basins to be investigated are chosen by the criterion that their minima differ in structure most from those of the initial basin and

other basins that have been sampled already. Alternatively, one may be interested in the dynamics of the system, so that preference is given to basins with low energy transition states. In the present study we mainly employ the first strategy that allows the system to pass into every new basin that has not been sampled previously since we are making an essentially complete survey of the all-*trans* PES. For finding the low energy regions pertaining to the various *cis* isomers, we choose instead to go to basins of minimal energy because the *cis* isomers have relatively low energies but are separated by high barriers.

The LD simulations were performed with the CHARMM program [10], using the polar hydrogen parameter set for peptides and proteins (param19) [15] and the ACS implicit solvation model [11]. The friction coefficient in the Langevin equations was set equal to 64 ps^{-1} [16] and a time step of 1 fs was used. For quenching, a combination of steepest descent (50 steps) and the adopted-basis set Newton-Raphson minimization methods (usually 300 steps) were employed [10]. Saddles were found with the TRAVEL algorithm [17].

The first survey of the PES was performed at 500 K with a LD simulation of 50 ns. A total of 408 minima and 4800 transition states were found by successive confinement with barriers up to 23 kcal/mol, relative to the global minimum. By contrast, only 97 minima and 670 transition states with barriers up to 8 kcal/mol were obtained with a conventional LD simulation of the same length; for minima with energies below -121 kcal/mol, the conventional LD sampling was relatively complete, but very few minima (21 versus 344 by successive confinement) were found above that value. No *cis* isomers were found during conventional LD, while several *cis* isomers (1,2,4,6; see Fig. 2 for the definition) were found by the confinement method. To obtain more efficient sampling of all the *cis* isomers, another LD trajectory of 6.4 ns was calculated at a temperature of 800 K.

Figure 2 presents a coarse grained disconnectivity graph for the entire PES, including all 16 isomers. Since the barriers that are crossed go up to 25 kcal/mol relative to the global minimum (27 kT for $T = 500$ K), it is of interest to illustrate how the confinement simulation progresses in finding the *cis* isomers. In going from the all-*trans* minimum energy basin (0) to mono *cis* (2), the ladder followed at 500 K is -126.1100, (-121.4047), -121.6296, (-113.3441), -122.4638, where the numbers are the energies of the minima and the numbers in parentheses are those of the barriers between them (all in kcal/mol); from all-*trans* (0) to mono-*cis* (4), the ladder is: -125.0197, (-118.9889), -121.7067, (-120.0868), -123.2215, (-110.9361), -123.6807. From all-*trans* to all-*cis* at 800 K, the ladder is 0 to 4: -124.4033, (-109.9052), -123.4851; from 4 to 6: -118.8571, (-106.3595) -121.2710; from 6 to 14: -113.5091, (-97.3553), -111.8034; and from 14 to 15: -120.6720, (-102.2448), -115.8970. Although the barriers crossed in the ladders shown for illustration go as high as 18.5 kcal/mol, a detailed analysis of the surface shows that no barriers higher than 11 to 12 kcal/mol have to be crossed to cover the entire PES, including the *cis* isomers.

As can be seen from Fig. 2, the all-*trans* isomer (0) contains the conformers of lowest energy, including the α-helical turn (α), which is the absolute minimum, and an extended β-like strand (β), which is the second lowest conformer. The right panel shows all of the all-*trans* conformers that are connected by barriers lower than -120 kcal/mol. It is clear that there are many low energy minima and the surface is complex but not funnel-like. Comparison with Fig. 8a of Ref. [5] shows that the all-*trans* vacuum surface and the surface with implicit solvent are significantly different. In vacuum, both the α-helical and extended β-strands are higher energy minima; the lowest energy vacuum structures have two hydrogen bonds, in contrast to only one for the α-helical structure, since hydrogen bonds play a more important role in vacuum. Most significantly the shape of the low energy region is quite different. The vacuum principal component surface [18] is a rather flat well until an energy of about 5 kcal/mol above the minimum is reached, and then the surface separates into a number of deep (funnel-like) superbasins. By contrast, the solvated surface is very rugged to a higher energy than in the vacuum case and has no overall tendency to decrease in energy toward the global minimum. It resembles the low temperature effective energy surface found in the cubic lattice Monte Carlo folding simulations for a 27-bead heteropolymer protein model (see Fig. 3 of [19]), much more than a simple folding funnel. This could be significant for protein folding.

Once the PES has been explored, the confinement simulations can be used to determine the equilibrium properties of the basins of interest and the kinetics of transitions from one basin or superbasin to another. While the system is confined to a current basin (i), one can calculate the probability $Q_{ji}(\tau)$ that the system will be found in the jth basin (including the original basin i) at the subsequent quenching after a time interval τ. Successively confining the system to the basins of the PES, one obtains the transition probability matrix $\mathbf{Q}\{Q_{ij}(\tau)\}$ that is related to the reaction rate matrix $\mathbf{W}\{W_{ji}\}$ as $\mathbf{Q}(\tau) = \exp(\mathbf{W}\tau)$. The kinetics can be described directly in the terms of the transition probabilities, if we treat the transitions between the minima as a Markovian process in the discrete time domain of quenching intervals τ. The time evolution of the vector of states $\mathbf{P}(t) = \{P_j(t)\}$ (the probability for the system to be found in basin j at time t) after n successive τ steps obeys the equation $\mathbf{P}(t = n\tau) = \mathbf{Q}^n(\tau)\mathbf{P}(t = 0)$.

Confinement simulations to determine the kinetics were done for 3.6 ns at 300 K. Sixty-nine basins and 289 transition states connecting them were found; only basins with an energy less than -123 kcal/mol were included. The results are in accord with those obtained from the 500 K

TABLE I. Equilibrium residence probabilities. $U_{\min}$ is the minimum energy of the conformer (kcal/mol). n_{conv} is the absolute number of transitions into the given basin from any other basin observed in the conventional simulations. The superscripts "conf" and "conv" refer to confinement and conventional simulations.

$U_{\min}$	n_{conv}	$\mathbf{P}_{\text{eq}}^{\text{conf}}$	$\mathbf{P}_{\text{eq}}^{\text{conv}}$
−127.5923 [α]	20	0.008 90	0.006 54
−125.0197	77	0.001 43	0.001 58
−126.8182 [β]	6702	0.545 08	0.540 80
−125.8585	3098	0.041 65	0.040 52
−125.8013	3432	0.058 40	0.055 09
−124.8512	5260	0.005 67	0.006 05
−126.0183	8320	0.096 73	0.085 13
−125.9658	633	0.073 92	0.082 43

simulations for this energy range. In each of the basins, the system made 150 attempts to pass into other basins; a quenching interval of 1 ps was used. A conventional LD (LD_{conv}) run of 10 ns found approximately the same number of minima and transition states (53 and 217, respectively), because LD_{conv} adequately covers this low energy part of the PES (see above). The discrepancy between the corresponding elements of the transition matrices $\mathbf{Q}_{\text{conv}}$ and $\mathbf{Q}_{\text{conf}}$ are within the statistical error range. The main differences between the two matrices is that $\mathbf{Q}_{\text{conv}}$ gives very good estimates for the transition probabilities for the small number of low-lying conformers, whereas the $\mathbf{Q}_{\text{conf}}$ matrix gives reasonable estimates for all the investigated conformers.

The equilibrium residence probabilities $\mathbf{P}_{\text{eq}}$ can be calculated directly from the conventional simulation by counting the number of times the system was found in a particular basin and, indirectly, from the confinement simulation if the transition matrix $\mathbf{Q}$ is irreducible (i.e., there exists a path between all pairs of the states, direct or indirect), from the equation $\mathbf{P}_{\text{eq}} = \mathbf{Q}\mathbf{P}_{\text{eq}}$. The values of $\mathbf{P}_{\text{eq}}$ calculated in these ways for some of the low-lying conformers are listed in Table I. It is seen that as the number of events in the conventional simulation increases, the agreement between the residence probabilities calculated from the two methods improves. Since the sampling of the surface by the confinement approach was essentially uniform, it follows that any pronounced discrepancy in the residence probabilities (e.g., for the −127.5923 level for which there are only a few events) arises from statistical errors in the direct simulation. A striking result in Table I is that the lowest energy structure (the α-helix with energy −127.5923 kcal/mol) has a much smaller residence probability than the β-strand basin. This is due primarily to the difference in configurational entropy; at 300 K the average energy difference is −0.8 kcal/mol and the entropic contribution is −2.45 kcal/mol.

This report has demonstrated the power of the confinement method for a simple biologically interesting molecule, the solvated tetrapeptide. Applications to larger systems are in progress.

S. V. K. and S. F. Ch. acknowledge the support from Russian Foundation for Basic Research, Grant No. 9903-33299, and from the Department of Education of Russian Federation, the Programme "Basic Problems in the Natural Sciences." The Laboratoire de Chimie Biophysique is partly supported by the CNRS (ISIS CR83), and by the Ministraire d'Education Nationale.

[1] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, Biochemistry **14**, 5355 (1975).
[2] R. Elber and M. Karplus, Science **235**, 318 (1987).
[3] M. Ferrand, A. J. Dianoux, W. Petry, and G. Zaccai, Proc. Natl. Acad. Sci. U.S.A. **90**, 9668 (1993).
[4] A. R. Dinner, A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus, Trends Biochem. Sci. **25**, 331 (2000).
[5] O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
[6] D. Wales, M. A. Miller, and T. R. Walsh, Nature (London) **394**, 758 (1998).
[7] D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic Press, San Diego, 1996).
[8] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1989).
[9] D. J. Wales and H. A. Scheraga, Science **284**, 1368 (1999); A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997); P. Labastie and R. L. Whetten, Phys. Rev. Lett. **65**, 1567 (1990); C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995); C. Bartels and M. Karplus, J. Phys. Chem. B **102**, 865 (1998).
[10] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[11] M. Schaefer, C. Bartels, and M. Karplus, J. Mol. Biol. **284**, 835847 (1998).
[12] R. Czerminsky and R. Elber, J. Chem. Phys. **92**, 5580 (1990).
[13] S. F. Chekmarev and S. V. Krivov, Chem. Phys. Lett. **287**, 719 (1998); Phys. Rev. E **57**, 2445 (1998).
[14] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[15] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[16] P. E. Smith, B. M. Pettitt, and M. Karplus, J. Phys. Chem. **97**, 6907 (1993).
[17] S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).
[18] O. M. Becker, J. Mol. Struct. (Theochem.) **398**, 507 (1997).
[19] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).